Comparing and Identifying Optimal Healthcare Databases for Comparative Effectiveness Research

Anokhi J Kapasi, Sharmila A Kamani, Vidya Moorthy, **Bao Nguyen-Khoa**, Varinder P Singh, Earl L Goehring Jr, **Judith K Jones**
DGI, LLC, Arlington, VA, USA E-mail: info@bridgetodata.org

➢ Comparative effectiveness research (CER) involves a thorough understanding of optimal resources for evaluating and comparing health outcomes and clinical effectiveness of medical treatments or health services.

➢ This requires familiarity with epidemiology, patient populations in various settings, and treatment patterns.

➢ A powerful step in achieving quality CER is to know which healthcare databases are available and their strengths and limitations.

➢ B.R.I.D.G.E. TO DATA® (B.R.I.D.G.E.; www.bridgetodata.org) an international resource of database profiles, may serve as one resource for CER studies.

RESEARCH OBJECTIVES

To provide descriptive information on how researchers may:
(1) Identify multiple databases and their attributes suitable for a CER study, and
(2) Compare the structure and components of different healthcare databases, including those appropriate for CER studies.

STUDY DESIGN

Case Study: A CER analyst must determine whether there are differences in length of stay, health outcomes, re-admissions, and costs between **pediatric inpatients** treated by physicians that request laboratory testing in comparison to those that do not.

A search was conducted in B.R.I.D.G.E. to identify databases collecting necessary data for the proposed CER study using criteria shown in **Figure 1**.

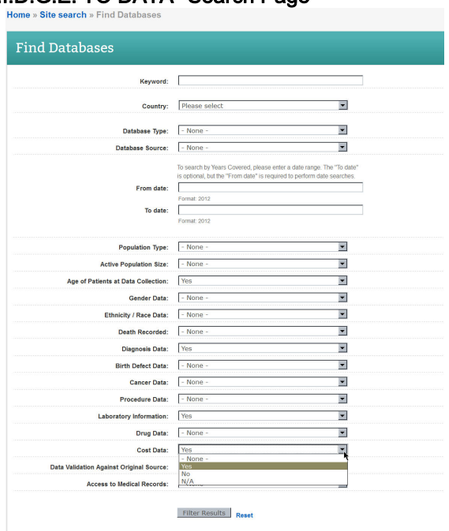
Figure 1. B.R.I.D.G.E. TO DATA® Search Page

Age = Yes

Diagnosis Data = Yes

Cost Data = Yes

Laboratory Data = Yes



A 100% match was identified either by a (i) 100% relevancy ranking OR (ii) manual review of profiles with a 75% match that was adjudicated using supplemental information in the profiles.

Search results were further narrowed by excluding databases that did not have adequate data on pediatric patients (<18 years), hospitalization, death, or procedures.

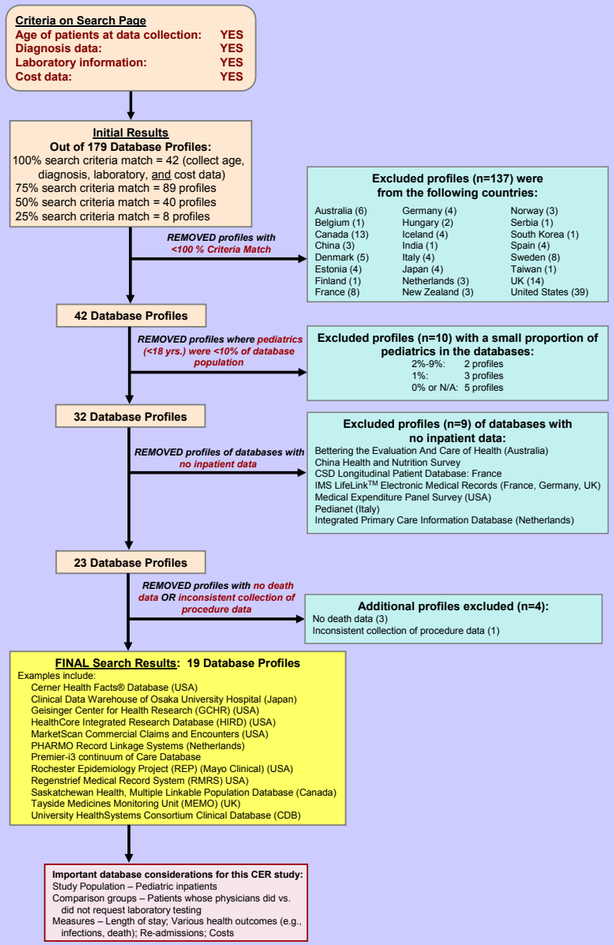
PRINCIPAL FINDINGS

The search yielded 179 profiles that matched at least one criterion (see **Figure 2**), while 42 profiles matched all four criteria. Profiles not matching all four search criteria were excluded from this analysis (n=137).

Profiles were further screened and excluded if:
- Patients aged <18 years accounted for <10% of the database (n=10);
- Database did not collect, or link to, inpatient data (n=9);
- Death information could not be obtained (n=3); or
- Procedures/services were inconsistently recorded (n=1).

Of the final 19 profiles from North America (15), Europe (3), and Asia (1), 11 collect medical claims data, and 6 collect electronic medical records.

Figure 2. Criteria-based search conducted in www.bridgetodata.org for CER case study (179 Database Profiles worldwide as of June 11, 2012)



Each of the 75 data fields used in structured profiles in B.R.I.D.G.E. can be compared side-by-side by a CER analyst to identify the most appropriate database(s) for evaluating whether there is a correlation between physician laboratory test order and health costs/outcomes for hospitalized pediatric patients (**Table 1**).

Table 1. Excerpt from B.R.I.D.G.E. TO DATA® comparing data elements in 3 selected databases that can be utilized in this proposed CER study. The database fields in red represent fields used as search/exclusion criteria.

FIELD NAMES	Cerner Health Facts® Database (USA)	MarketScan Medicaid Database (USA)	Tayside Medicines Monitoring Unit (MEMO) (UK)
Coordinating Country/ Region	USA Kansas City, MO (USA)/ Over 480 contributing facilities throughout USA	United States/ All regions of USA (The states are geographically dispersed)	United Kingdom/ Tayside, Scotland
Database Source	Electronic Health or Medical Record(s) (aka EHR/EMR) About 96% of the inpatient encounters have medication, general lab, or billing data; 70% of Emergency and 60% of Outpatient visits have activity in the 3 main subject areas.	Medical Insurance Claims	Electronic Medical Records
Years Covered	2000 - Present (January 1, 2000 - Present)	1988 - Present (With increasing number of data sources included)	1989 - Present
Population Type	Outpatient/Non-institutionalized Inpatient / Newborn Emergency Room (ER/ED) Population Other specialty institutions (Nursing homes, Children's hospitals, etc.) Visit types (Clinic, Urgent Care, Dialysis, Obstetrics; Day Surgery, etc.) Other (Billing, Community, Dental, Home Health, Hospice, Observation, Short stay / 24 hr stay, Occupational Health, etc.)	Insured (Medicaid recipients for several states)	General Population
Patient Type	Inpatient Outpatient Emergency Room (ER/ED) Outpatient data are comprised of two distinct populations: 1. Outpatient clinics associated with a hospital system: In this case, longitudinality exists between all venue types associated with that system. Therefore, a patient can be tracked across outpatient, hospital, and ER visits; 2. Physician Offices not associated with a hospital system: In this case, a patient can be tracked across multiple visits to the same physician office.	Inpatient and Outpatient Emergency Room (ER/ED)	Inpatient and Outpatient
Database Population Size (Range)	20 - 50 Million [As of January 2012, there were 35,001,010 unique patients and 156,198,274 encounters (acute admissions, emergency and ambulatory visits) in the database]	5 - 20 Million (16.5 Million)	1 - 5 Million
Approximate Percentage of Participants <18 years and those >65 years	<18 years = 15% >65 years = 40%	< 18 years = 60% > 65 years = 9%	< 18 = 18.9% > 65 = 19.3%
Death Recorded	Yes Information is recorded on in-hospital mortality and cause of death	Yes Death information is available on hospital discharge records	Yes
Diagnosis Data	Yes All diagnoses made at the time of visit / discharge and comorbidities recorded during the patient medical history are collected	Yes	Yes Each SMR1 record (Scottish Morbidity Record) has one principal and five other diagnostic fields coded according to the International Classification of Diseases 99/10th Revision
Diagnoses Coded	ICD-9 DRG Other (Primary diagnosis coding system: ICD-9. Major Diagnostic Categories (MDC) and DRG are secondary groupings applied to the ICD-9-CM codes for analysis purposes.)	ICD-9-CM DRG	ICD-9-CM ICD-10
Diagnoses: Maximum Number of Codes Allowed	9 (Principal diagnosis and up to 8 secondary diagnoses)	15 (Up to 15 DX on admission record)	6 (Main diagnosis plus 5 others)
Laboratory Information	Yes Data on time-stamped laboratory test orders and 1,888,833,604 lab test results (as of January 2012) are recorded. Laboratory tests include: Chemistry, Hematology, Urine analysis, Coagulation, Bloodbank, Anatomic pathology, Microbiology, Immunology-Serology, and Flow cytometry. Specific laboratory data include: result, unit of measurement, dates & times for lab (ordered, received, etc.); medical specialty of the ordering physician; type of clinical care provided at location. Microbiology findings include the organism name for positive findings, type of result, specimen, and microbiology susceptibility.	No	Yes Clinical chemistry and cancer registration data are available
Drug Data	Yes: Prescription & OTC Other [There are data on drugs, vaccines, and devices, all time-stamped to the minute. There are ~8,000 drugs (by name and brand) from pharmacy orders dispensed by pharmacy. The database does not include medications managed by other departments (e.g., Surgery, Radiology).]	Yes: Prescription only (Filled prescriptions (retail or mail order), physician administered - specialty, immunizations/vaccines, etc. J-codes found in service level files.)	Yes: Prescription only Primary Care Databases - Drug exposure is collected from dispensed prescriptions in the primary care. The dispensed prescribing database in MEMO contains patient specific information from over 11 million prescriptions dispensed in Tayside since January 1989 (e.g., prescription date, prescriber, dosage, prescribed amount).
Cost Denomination	US Dollars	US Dollars	British Sterling/Pounds
Type of Cost Data	Yes Hospital charges from the UB-92 and CMS1500 billing are captured. This database does not contain revenue codes or detail charge master data. Charge data can be converted to Costs through methodologies developed by Cerner.	Yes Several types, including Average Wholesale Price, Actual, and Paid	Yes Actual
Database Contact Data	Daniel Aguilar, MPH, MBA Account Executive Cerner LifeSciences USA Phone: 310-598-4533 (Direct) Phone: 816-309-5771 (Mobile) E-Fax: 816-936-1933 Email: Daguilar@cerner.com	Stella Chang, MPH Director, Information Assets Thomson Reuters 4301 Connecticut Ave. NW, Suite 330 Washington, DC 20008 USA Phone: +1 (703) 938-2351 Mobile: +1 (202) 257-8408 Email: stella.chang@thomsonreuters.com	Prof Tom MacDonald Hypertension Research Centre & Medicines Monitoring Unit University of Dundee Ninewells Hospital & Medical School Dundee DD1 9SY UNITED KINGDOM Phone: +44 1382 632852 Email: tom@memo.dundee.ac.uk

CONCLUSIONS

Through this case study we have demonstrated how B.R.I.D.G.E. provides an approach that supports decision-making for CER, serves as a useful tool to identify and compare health database attributes, and can also be used as a teaching tool for understanding healthcare databases.

IMPLICATIONS: Policy, Delivery or Practice

✓ Understanding the types of data collected by population healthcare databases, and their strengths and limitations, are part of the core competency for CER.

✓ The B.RI.D.G.E. resource can support policy advisors, physicians, and healthcare professionals in conducting public health research, or for communicating and comprehending CER findings by providing structured information on epidemiologic databases.

RESEARCH FUNDER: This study was self-funded by DGI, LLC.

AcademyHealth 2012 Annual Research Meeting, Orlando, FL, USA. June 25, 2012.

Comparing and Identifying Optimal Healthcare Databases for Comparative Effectiveness Research

Anokhi J Kapasi, Sharmila A Kamani, Vidya Moorthy, Bao Nguyen-Khoa,
Varinder P. Singh, Earl L Goehring Jr, Judith K Jones

DGI, LLC, Arlington, VA, USA

Presented at the AcademyHealth 2012 Annual Research Meeting, Orlando, FL, USA. June 25, 2012.

DGI Center for Health Research and Education, LLC
1616 N Fort Myer Drive, Suite 1430 • Arlington, VA 22209 USA
www.bridgetodata.org T: +1 703 894 0804 info@bridgetodata.org